

日本国特許

PATENT OFFICE
JAPANESE GOVERNMENT

REC'D 03 MAR 2000

WIPO

PCT

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出願年月日
Date of Application:

1999年 2月16日

出願番号
Application Number:

平成11年特許願第037559号

出願人
Applicant(s):

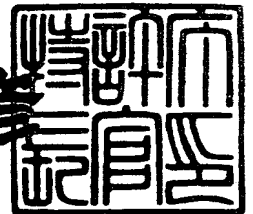
有限会社ジーエムアンドエム

PRIORITY DOCUMENT
SUBMITTED OR TRANSMITTED IN
COMPLIANCE WITH
RULE 17.1(a) OR (b)

2000年 1月28日

特許庁長官
Commissioner,
Patent Office

近藤 隆彦



出証番号 出証特2000-3002492

【書類名】 特許願

【整理番号】 98P760T03

【提出日】 平成11年 2月16日

【あて先】 特許庁長官 伊佐山 建志 殿

【国際特許分類】 H04R 25/00

【発明者】

 【住所又は居所】 東京都杉並区善福寺2丁目8番3号

 【氏名】 大場 俊彦

【特許出願人】

 【住所又は居所】 東京都港区高輪4丁目11番24号

 【氏名又は名称】 有限会社ジーエムアンドエム

 【代表者】 大場 五美

【代理人】

 【識別番号】 100067736

 【弁理士】

 【氏名又は名称】 小池 晃

【選任した代理人】

 【識別番号】 100086335

 【弁理士】

 【氏名又は名称】 田村 榮一

【選任した代理人】

 【識別番号】 100096677

 【弁理士】

 【氏名又は名称】 伊賀 誠司

【手数料の表示】

 【予納台帳番号】 019530

 【納付金額】 21,000円

【提出物件の目録】

 【物件名】 明細書 1

【物件名】	図面	1
【物件名】	要約書	1
【プルーフの要否】	要	

【書類名】 明細書

【発明の名称】 音声生成装置及び方法

【特許請求の範囲】

【請求項 1】 音声言語障害を有して発せられた音声を検出して音声信号を生成する音響電気変換手段と、

上記音響電気変換手段からの音声信号に基づいて音声認識をする処理を行う音声認識手段と、

音声言語障害を有しないで発せられた音声を予めサンプリングすることで生成した音声データを記憶する記憶手段と、

上記音声認識手段からの認識結果に基づいて、上記記憶手段に記憶された音声データを用いて出力する音声を示す音声情報を生成する音声情報生成手段と、

上記音声情報生成手段からの音声情報を用いて音声を出力する音声出力手段とを備えることを特徴とする音声生成装置。

【請求項 2】 上記音声認識手段からの認識結果を用いて音声出力手段から出力した音声の内容を表示する表示手段を備えること

を特徴とする請求項 1 記載の音声生成装置。

【請求項 3】 通信回線を通じて音声及び/又は音声データを入力して上記音響電気変換手段及び/又は音声認識手段に入力するとともに上記出力手段からの認識結果を通信回線に出力する上記通信手段を備えること

を特徴とする請求項 1 記載の音声生成装置。

【請求項 4】 上記音響電気変換手段は、音声言語障害を有して発せられた音声として、音声言語障害を是正するために用いられる補助的手段や装置を用いて発せられた音声を検出して音声信号を生成すること

を特徴とする請求項 1 記載の音声生成装置。

【請求項 5】 上記音響電気変換手段は、音声言語障害を有して発せられた音声として、音声言語障害を是正するために用いられる、喉頭摘出後になされる代用発声法を用いて発せられた音声を検出して音声信号を生成すること

を特徴とする請求項 4 記載の音声生成装置。

【請求項 6】 上記音響電気変換手段は、音声言語障害を有して発せられた音声として、音声言語障害を是正するために用いられる、喉頭摘出後になされる代用発声法を用いない音声言語障害者から発せられた音声を検出して音声信号を生成すること

を特徴とする請求項 4 記載の音声生成装置。

【請求項 7】 画像を撮像する撮像手段を備え、

上記撮像手段は、撮像した画像を少なくとも表示手段に出力すること

を特徴とする請求項 2 記載の音声生成装置。

【請求項 8】 上記撮像手段は、使用者の視力に基づいて、撮像した画像について画像変換処理を施して表示手段に出力すること

を特徴とする請求項 7 記載の音声生成装置。

【請求項 9】 少なくとも上記撮像手段は、使用者に対して着脱自在となされていること

を特徴とする請求項 7 記載の音声生成装置。

【請求項 10】 通信回線を通じて画像を入力するとともに上記音声情報生成手段からの認識結果を通信回線に出力する上記通信手段を備えること

を特徴とする請求項 1 記載の音声生成装置。

【請求項 11】 少なくとも上記音声出力手段は、使用者に対して着脱自在となされていること

を特徴とする請求項 1 記載の音声生成装置。

【請求項 12】 音響電気変換手段、音声認識手段、記憶手段、音声情報生成手段及び音声出力手段を分割して複数の装置とし、各装置間をワイヤレスとして少なくとも音声情報の送受信を行うこと

を特徴とする請求項 1 記載の音声生成装置。

【請求項 13】 音声言語障害を有しないで発した音声をサンプリングし、

音声言語障害を有して発せられた音声を検出して音声信号を生成し、

上記音声信号に基づいて音声認識をして認識結果を生成し、

上記認識結果に基づいて、音声言語障害を有しないで発せられた音声をサンプリングすることで生成した音声データを用いて出力する音声を示す音声情報を生

成し、

上記音声情報を用いて音声を出力すること
を特徴とする音声生成方法。

【請求項 14】 上記認識結果を用いて出力した音声の内容を表示すること
を特徴とする請求項 13 記載の音声生成方法。

【請求項 15】 通信回線を通じて音声又は音声データを入力し、
音声データを用いて言語認識をして認識結果を生成すること
を特徴とする請求項 13 記載の音声生成方法。

【請求項 16】 音声言語障害を有して発せられた音声として、音声言語障害
を是正するために用いられる補助的手段や装置を用いて発せられた音声を検出し
て音声信号を生成すること
を特徴とする請求項 13 記載の音声生成方法。

【請求項 17】 音声言語障害を有して発せられた音声として、音声言語障害
を是正するために用いられる喉頭摘出後の代用発声法を用いて発せられた音声を
検出して音声信号を生成すること
を特徴とする請求項 16 項記載の音声生成方法。

【請求項 18】 音声言語障害を有して発せられた音声として、音声言語障害
を是正するために用いられる代用発声法を用いない音声言語障害者から発せられ
た音声を検出して音声信号を生成すること
を特徴とする請求項 16 記載の音声生成方法。

【請求項 19】 画像を撮像し、
撮像した画像を表示すること

を特徴とする請求項 13 項記載の音声生成方法。

【請求項 20】 使用者の視力に基づいて、撮像した画像について画像変換処
理を施して表示すること
を特徴とする請求項 19 記載の音声生成方法。

【請求項 21】 通信回線を通じて画像を入力するとともに上記出力手段から
の音声情報を通信回線に出力すること
を特徴とする請求項 13 記載の音声生成方法。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明は、音声言語障害を持つ者より発せられた音声や音声言語障害を是正するために用いる補助的装置や手段（例として喉頭摘出後の代用発声法（speech production substitutes））により発せられた音声を変換して出力する音声生成装置及び方法に関する。

【0002】

【従来の技術】

従来において、例えば喉頭摘出して音声障害をもつ人は、通常の声帯振動による発声機構を失い、音声生成が困難になる。

【0003】

現在まで、喉頭摘出後の代用発声法として、音源としての振動体の性質から大別するとゴム膜（笛式人工喉頭）やブザー（電気人工喉頭（経皮型、埋込み型））等の人工材料を用いる方法と、下咽頭や食道粘膜を使用する方法（食道発声、気管食道瘻発声、ボイスプロステシス（voice prostheses）使用の気管食道瘻発声）がある。また、その他の代用発声法としては、口唇を動かしたときに生じる筋電図を利用したものや聴力障害による発声障害者のために種々の音声処理技術を利用した発声発話訓練装置、パラトグラフ（palatograph）によるものや口腔内の振動子によるものが報告されている。

【0004】

【発明が解決しようとする課題】

上記代用発声法に共通してみられるのは、喉頭摘出前の本人自身の本来の正常な状態での声帯振動によるものではないので、生成する音声の音質が悪くなく、本来正常であった本人が発していた声とはかけ離れているという問題点が挙げられる。

【0005】

そこで、本発明は、上述したような実情に鑑みて提案されたものであり、喉頭摘出や舌口腔底切除や構音障害等による音声言語障害を有する人達が本来自身が

もつ、或いは自在に変換させて自然な音声で発声することを可能とする音声生成装置及び方法を提供することを目的とする。

【0006】

【課題を解決するための手段】

上述の課題を解決する本発明に係る音声生成装置は、音声言語障害を有して発せられた音声を検出して音声信号を生成する音響電気変換手段と、音響電気変換手段からの音声信号に基づいて音声認識をする処理を行う音声認識手段と、音声言語障害を有しないで発せられた音声や希望する音声を予めサンプリングすることで生成した音声データを記憶する記憶手段と、音声認識手段からの認識結果に基づいて、記憶手段に記憶された音声データを用いて出力する音声を示す音声情報を生成する音声情報生成手段と、音声情報生成手段からの音声情報を用いて音声を出力する音声出力手段とを備えることを特徴とするものである。

【0007】

また、本発明に係る音声生成方法は、音声言語障害を有しないで発した音声や希望する音声をサンプリングし、音声言語障害を有して発せられた音声を検出して音声信号を生成し、音声信号に基づいて音声認識をして認識結果を生成し、認識結果に基づいて、音声言語障害を有しないで発せられた音声をサンプリングすることで生成した音声データを用いて出力する音声を示す音声情報を生成し、音声情報を用いて音声を出力することを特徴とする。

【0008】

【発明の実施の形態】

以下、本発明の実施の形態について図面を参照しながら詳細に説明する。

【0009】

本発明は、例えば図1及び図2に示すように構成された音声生成装置1に適用される。この音声生成装置1は、図1に示すように、ヘッドマウントディスプレイ (head mounted display: HMD) 2と、音声認識、音声情報の生成等を行うコンピュータ部3との間を光ファイバーケーブル4で接続してなる携帯型のものである。また、コンピュータ部3は、例えば使用者の腰部に装着されるような支持部5に付属して配設されており、当該支持部5に付属したバッテリー6から電力が供

給されることで駆動するとともに、HMD 2 を駆動させる。

【 0 0 1 0 】

HMD 2 は、使用者の目前に配置されるディスプレイ部 7 と、使用者からの音声を検出する使用者用マイクロホン 8 と、使用者に音声を出力する音声出力部 9 と、使用者の頭部に上述の各部を配置させるように支持する支持部 1 0 とを備える。

【 0 0 1 1 】

ディスプレイ部 7 は、使用者の目前に配されることで例えば使用者用マイクロホン 8 及び／又は後述の外部用マイクロホン 1 1 で検出した音声の意味内容を表示する。なお、この表示部 2 6 は、コンピュータ部 3 からの命令に応じて、上述の音声の意味内容のみならず、他の情報を表示しても良い。

【 0 0 1 2 】

使用者用マイクロホン 8 は、位置を問わず、使用者の操作に応じて無指向性マイク、単一指向性（超指向性等）マイク、双指向性マイク、ダイナミックマイク、コンデンサーマイク（エレクトレットマイク）、ズームマイク、ステレオマイク、MSステレオマイク、ワイヤレスマイク等を用いても良い。

【 0 0 1 3 】

更に、使用者用マイクロホン 8 及び外部用マイクロホン 1 1 は、別個に設ける一例のみならず、一体に構成されたものであっても良い。

【 0 0 1 4 】

支持部 1 0 は、例えば形状記憶合金等の弾性材料等からなり、使用者の頭部に固定可能とすることで、上述のディスプレイ部 7、使用者用マイクロホン 8、音声出力部 9 を所定の位置に配設可能とする。なお、この図 1 に示した支持部 1 0 は、使用者の額から後頭部に亘って支持部材を配設することでディスプレイ部 7 等を所定位置に配設するものの一例について説明したが、所謂ヘッドホン型の支持部であっても良いことは勿論であり、音声出力部 9 を両耳について設けても良い。

【 0 0 1 5 】

コンピュータ部 3 は、例えば使用者の腰部に装着される支持部 5 に付属されて

なる。このコンピュータ部 3 は、図 2 に示すように、例えばマイクロホンで検出して生成した電気信号が入力される。このコンピュータ部 3 は、電気信号を処理するためのプログラムを格納した記録媒体、この記録媒体に格納されたプログラムに従って音声認識、音声情報の生成処理を行う CPU (Central Processing Unit) 等を備えてなる。

【0016】

つぎに、本発明を適用した音声生成装置 1 の電氣的な構成について図 2 を用いて説明する。この音声生成装置 1 は、音声を検出して音声信号を生成する上述のマイクロホン 8、11 に相当するマイクロホン 21 と、マイクロホン 21 で生成された音声信号が入力され音声認識処理を行う上述のコンピュータ部 3 に含まれる認識部 22、認識部 22 からの認識結果に基づいて音声情報を生成する上述のコンピュータ部 3 に含まれる音声情報生成部 23 と、音声データが記憶され認識部 22 及び音声情報生成部 23 にその内容が読み込まれる上述のコンピュータ部 3 に含まれる記憶部 24 と、音声情報生成部 23 からの音声情報を用いて音声を出力する上述の音声出力部 9 に相当するスピーカ部 25 と、音声情報生成部 23 からの音声情報を用いて当該音声情報が示す内容を表示する上述のディスプレイ部 7 に相当する表示部 26 とを備える。

【0017】

上記マイクロホン 21 は、例えば喉頭摘出後の代用発声法を用いて発せられた使用者からの音声を検出して、当該音声に基づく音声信号を生成する。そして、このマイクロホン 21 は、生成した音声信号を認識部 22 に出力する。使用者の口元付近に配設され、使用者が発した音声を検出する。そして、この使用者用マイクロホン 8 は、使用者からの音声を電気信号に変換してコンピュータ部 3 に出力する。

【0018】

音声出力部 9 は、例えば側面に設けられ外部からの音声を検出する外部用マイクロホン 11 を備えている。この音声出力部 9 は、外部用マイクロホン 11 により使用者との話し相手の音声を検出することで電気信号を生成させ、外部用マイクロホン 11 から信号処理回路に生成した電気信号を出力する。

【0019】

上記代用発声法としては、例えば人工喉頭（電気式、笛式）、食道発声及び種々の音声再建術をを実現するための機構である。

【0020】

上記認識部22は、マイクロホン21からの音声信号を用いて音声認識処理を行う。この認識部22は、例えば内部に備えられたメモリに格納した音声認識処理を行うためのプログラムに従った処理を行うことにより音声認識処理を実行する。具体的には、この認識部22は、記憶部24に格納された使用者の音声を示す音声データを参照し、マイクロホン21からの音声信号を言語として認識する処理を行う。この結果、この認識部22は、マイクロホン21からの音声信号に応じて認識結果を生成する。

【0021】

この認識部22は、例えば認識対象音声による分類と対象話者による分類の音声認識処理があり、認識対象音声による分類の音声認識処理では単語音声認識(isolated word recognition)と連続音声認識(continuous speech recognition)がある。また、音声情報生成部23は、連続単語音声認識には連続単語音声認識(continuous word recognition)と文音声認識(sentence speech recognition)、会話音声認識(conversational speech recognition)、音声理解(speech understanding)がある。また対象話者による分類では不特定話者型(speaker independent)、特定話者型(speaker dependent)、話者適応型(speaker adaptive)等がある。以上の音声認識を使用者の身体状態、利用状態及び使用目的に応じて行う。

【0022】

また、この認識部22は、マイクロホン21に入力した音声を記憶し、学習する機能を有する。これにより、認識部22は、更に認識率を向上させる。更に、この認識部22は、学習機能を備えることで出力する結果を正確にすることができる。

【0023】

上記記憶部24には、上記認識部22が入力された音声を認識するときに、入力された音声を検出することで生成した音声波形と比較される音声モデルデータ

が格納されている。また、記憶部 2 4 には、例えば喉頭摘出前に発声した声帯振動による発声機構を持つ使用者の音声や、出力することを希望する音声を予めサンプリングして音声データとして格納されている。

【 0 0 2 4 】

音声情報生成部 2 3 は、認識部 2 2 からの認識結果及び記憶部 2 4 に格納された使用者の音声を示す音声データを用いて、音声情報を生成する。このとき音声情報生成部 2 3 は、認識結果に応じて、記憶部 2 4 に格納された音声データを組み合わせることで音声情報を生成する。具体的には、この音声情報生成部 2 3 は、認識結果を用いて音声から音声分析をし、当該音声分析した音声の内容に応じて、音声データを再構成するという処理を行うことで、音声を示す音声情報を生成する。そして、音声情報生成部 2 3 は、生成した音声情報をスピーカ部 2 5 及び表示部 2 6 に出力する。

【 0 0 2 5 】

スピーカ部 2 5 は、上記音声情報生成部 2 3 で生成した音声を出力する。このスピーカ部 2 5 としては、例えば使用者から話し手に対して音声を出力するものであっても良く、更には、使用者の耳に対して発声するように音声を出力するものであっても良い。また、使用者の耳に対して発声するように音声を出力するスピーカ部 2 5 は、スピーカユニットの変換方式としてダイナミック型や静電型（コンデンサ型、エレクトロスタティック型）によるものでも良く、形状としてはヘッドフォン（オープンエア型、クローズド型、カナルタイプ等のイン・ザ・イヤ型等）によるものでも良い。また、スピーカ部 1 3 c は、従来の補聴器、集音器のスピーカによるものでも良く、使用者から話者に対して音声を出力するスピーカ部 2 5 は従来から用いられているスピーカ装置でよい。

【 0 0 2 6 】

表示部 2 6 は、上記音声情報生成部 2 3 で生成した音声情報が示す音声を表示する。

【 0 0 2 7 】

また、この音声生成装置 1 は、外部の通信ネットワークと接続された通信回路 2 7 を備えている。この通信回路 2 7 は、電話、携帯電話、インターネットや無

線、衛星通信等の通信ネットワークを介して例えば音声言語障害を有する者から発せられた音声が入力され、当該音声を認識部 22 に入力する。

【0028】

このように構成された音声生成装置 1 は、例えば喉頭摘出後に電気式人工喉頭を使って発声された音声マイクロホン 21 に入力された場合であっても、認識部 22 で音声認識し、記憶部 24 に格納された喉頭摘出前にサンプリングした音声を示す音声データを用いて音声情報生成部 23 で出力する音声を示す音声情報を生成するので、スピーカ部 25 から喉頭摘出前の使用者の音声を出力することができる。

【0029】

なお、上述した本発明を適用した音声生成装置 1 の説明においては、マイクロホン 21 で検出される喉頭摘出した人の音声である一例について説明したが、聴力障害による言語障害の一つである構音障害 (articulation disorders) を持つ者からの音声を検出したときであっても適用可能である。このとき、音声生成装置 1 は、言語障害の音声を音声データとして記憶部 24 に記憶しておき、当該発声者が発声したことに応じて記憶部 24 に格納された発声者の音声を示す音声データを参照して認識部 22 で音声認識処理を行い、音声情報生成部 23 で認識結果に応じて音声データを組み合わせることで音声情報を生成する処理を行うことにより、スピーカ部 25 から音声言語障害のない音声を出力するとともに、表示部 26 により音声情報に基づいた音声内容を表示することができる。

【0030】

したがってこの音声生成装置 1 によれば、例えば喉頭摘出者が代用発声法により発生した音声を表示部 26 に表示することで不自然な音声を訂正させることができる。

【0031】

更に、音声生成装置 1 は、例えば聴力障害による構音障害を持つ者は発声のためのフィードバックが得られず、「きょうは（今日は）」という音声「きょんわあ」となってしまふのを上述した処理を行うことにより正常な「きょうは（今日は）」という音声に訂正してスピーカ部 25 から出力することができる。

【0032】

更に、この音声生成装置1は、表示部26を備えているので、発声者の音声をスピーカ部25から正常な音声にして出力するとともに、発声者の音声内容を表示することにより難聴者の言語訓練にとって好適なシステムを提供することができる。

【0033】

つぎに、上述の音声情報生成部23が認識部22からの認識結果を用いて合成した音声情報を、変換させるときの種々の例について述べる。なお、変換処理の種々の例は、以下に述べる例に限定するものではない。

【0034】

この音声情報生成部23は、例えばアクセント等の処理を行うように変換処理を行って音声情報を出力しても良い。このとき、音声情報生成部23は、必要に応じて、特定の発音についてはアクセントの強弱を変化させるように音声情報を変換して出力するようにする。

【0035】

音声情報生成部23は、音声データを合成するとき、どのような内容の音声でも合成するときには規則による音声合成、滑らかな音声を合成するために可変長単位を用いた音声合成、自然な音声を合成するための韻律制御、また音声の個性付与のために音質変換を行って音声情報を生成しても良い。これは、例えば書籍「自動翻訳電話」ATR国際電気通信基礎技術研究所編 pp.177-209, 1994」に記載されている技術を適用することにより実現可能である。

【0036】

また、ボコーダ(vocoder)処理を用いても高品質の音声を合成することが可能である。例えば音声分析変換合成法STRAIGHT (speech transformation and representation based on adaptive interpolation of weighted spectrogram)等を施すことで実現可能である。

【0037】

更に、この音声情報生成部23は、文字情報から音声を作り出す音声合成(text to speech synthesis)技術を用いることにより話の内容に関する情報(音韻

性情報)や音の高さや大きさに関する情報(韻律情報)を聴力障害者の難聴の特性に合わせてその人の最も聞き易い音の高さに調整することも可能であり、他に話速変換技術(voice speed converting)、周波数圧縮(frequency compress)処理などの音声特徴量の変換処理を行う。また出力する音声の帯域を調整する帯域拡張(frequency band expansion)処理や、音声強調(speech enhancement)処理等を音声情報に施す。帯域拡張処理、音声強調処理としては、例えば「M. Abe, "Speech Modification Methods for Fundamental Frequency, Duration and Speaker Individuality," TECHNICAL REPORT OF IEICE, SP93-137, pp.69-75, 1994」にて示されている技術を用いることで実現可能である。

【0038】

更にまた、上記音声情報生成部23は、認識結果を用いて音声から言語を理解し、当該理解した言語を用いて音声データから音声情報を構成するという処理を行うのみならず、他の処理を認識結果に基づいて理解した言語を必要に応じて変換する処理を行っても良い。すなわち、この音声情報生成部23は、音声情報を構成するとともに、音声情報としてスピーカ部25に出力するときの速度を変化させる話速変換処理を行っても良い。すなわち、この話速変換処理は、使用者の状態に応じて適当な話速を選択することによりなされる。

【0039】

更にまた、この音声情報生成部23は、認識結果に応じて、例えば日本語の音声情報を英語の音声情報に変換して出力するような翻訳処理を行って出力しても良く、通信機能と合わせて自動翻訳電話にも応用可能である。更には音声情報生成部23は自動要約(automatic abstracting)を行い、「United States of America」を「USA」と要約するように変換して音声情報を出力しても良い。

【0040】

この音声情報生成部23行う自動要約処理としては、例えば文章内から要約に役立ちそうな手がかり表現を拾い出し、それらをもとに読解可能な文表現を生成する生成派の処理(文献「K. McKeown, D. R. Radev, "Generating Summaries of Multiple News Articles," In Proceedings of the Fourteenth Annual International ACM SIGIR Conference on Research and Development in Information

Retrieval, pp.68-73, 1995」、文献「E. Hovy, "Automated Discourse Generation using Discourse Structure Relations," Artificial Intelligence, 63, pp.341-385, 1993」参照)、要約を「切り抜き」と考えて処理し客観的評価が可能となるように問題を設定しようという立場の抽出派の処理(文献「J. Kupiec, et al., "A Trainable Document Summarizer," In Proceedings of the Fourteenth Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp.68-73, 1995」、文献「S. Miike, et al., "A Full-text Retrieval System with a Dynamic Abstract Generation Function," Proceedings of the seventeenth Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp.152-159, 1994」及び文献「H.P. Edmundson, "New Method in Automatic Abstracting," Journal of the ACM, 16, pp.264-285, 1969」参照)がある。更に、この音声情報生成部23は、例えば文献「M. Nakazawa, et al. "Text summary generation system from spontaneous speech," 日本音響学会講演論文集 1-6-1, pp.1-2, 1998」に記載されている手法(Partial Matching Method (PMM) と Incremental Reference Interval-Free連続DP(IRIFCDP)を用いて重要キーワードの抽出を行い、Incremental Path Method(IPM)を用いて単語認識を行う)を用いることが可能である。

【0041】

更にまた、この音声情報生成部23は、認識結果に応じて、特定の音素、母音及び子音、アクセント等において、消去したり、音声を出力することに代えてブザー音、あくび音、せき音、単調な音などを音声情報とともに出力するように制御しても良い。このとき、音声情報生成部23は、例えば文献「R. M. Warren, "Perceptual Restoration of Missing Speech Sounds," Science vol.167 pp. 392, 1970」や文献「R. M. Warren, C. J. Obusek, "Speech perception and phonemic restoration," Perception and psychophysics vol.9 pp.358, 1971」に記載されている手法を実現した処理を音声情報について行う。

【0042】

更にまた、音声情報生成部23は、認識結果を用いてホーン調となるように音

質を変換させて音声情報を出力しても良い。上記ホーン調とは、集音管を使ったもので、約2000Hz以下の帯域の音声を増幅させて、利得が約15dB程度とすることである。すなわち、このホーン調とは、管共鳴を用いた重低音を再生する技術により出力される音質である。この音声情報生成部23は、例えばU.S. PATENT No.4628528により公知となされているアコースティックウェーブ・ガイド (acoustic wave guide) 技術を用いて出力される音質に近似した音に変換して音声情報を出力する。ここで、音声情報生成部23は、例えば低音のみを通過させるフィルター処理を行って音声情報を出力する処理を行っても良く、例えばSUVAG (Systeme Universel Verbo-tonal d'Audition-Guberina) 機器を用いることにより、所定の周波数帯域の音声のみを通過させる種々のフィルタ処理を行って音声情報を出力する処理を行っても良い。

【0043】

更にまた、この音声情報生成部23は、例えばマイクロホン21に音楽が入力されたと判断したときには、音声情報を変換して表示部26に音符や色を表示するように処理を行っても良い。また、この音声情報生成部23は、音声のリズムなどが分かるために変換した音声のリズムを信号が点滅するように音声情報を変換して表示部26に表示しても良い。

【0044】

更にまた、この音声情報生成部23は、例えば警報等の発信音がマイクロホン21に入力されたと判断したときには、音声情報を変換することで表示部26に警報等がマイクロホン21で検出された旨の表示を行ったり、スピーカ部25に警報の内容を知らせるような内容を出力しても良く、例えば救急車のサイレンを聞いたら表示するだけでなく大音量で「救急車ですよ」とスピーカ部25から出力する。

【0045】

更にまた、音声情報生成部23は、過去に行った加工変換合成処理について記憶する機能を備えていても良い。これにより、音声情報生成部23は、過去に行った加工変換合成処理の改良を自動的に行う学習処理を行うことができ、加工変換合成処理の処理効率を向上させることができる。

【0046】

更にまた、この認識部22及び音声情報生成部23は、話し手の音声のみについての認識結果を生成して音声情報を生成し、スピーカ部25及び/又はディスプレイ部7に提示することで使用者に知らせる一例のみならず、例えば特定の雑音に対してのみ音声認識を行っても良い。要するに、認識部22及び音声情報生成部23は、入力した音について音声認識処理を行って、認識結果を使用者の身体状態、利用状態及び使用目的に応じて変換することで使用者が理解し易い表現で音声情報を生成して出力する処理を行う。

【0047】

更にまた、上述した本発明を適用した音声生成装置1の説明においては、記憶部24に予めサンプリングして格納した音声データを音声情報生成部23により組み合わせることにより音声情報を生成して出力するものの一例について説明したが、上記音声情報生成部23は、記憶部24に記憶された音声データを組み合わせる音声情報を生成するときに格納された音声データに変換処理を施す音声データ変換部を備えていても良い。このような音声データ変換部を備えた音声生成装置1は、例えばスピーカ部25から出力する音声の音質を変化させることができる。

【0048】

更にまた、上述した本発明を適用した音声生成装置1の説明においては、例えば喉頭摘出前の使用者の音声を予めサンプリングすることにより得た音声データを記憶部24に格納するものの一例について説明したが、記憶部24には、一つの音声データのみならず複数の音声データを予めサンプリングして格納しても良い。すなわち記憶部24には、例えば喉頭摘出前に発せられた音声を予めサンプリングした音声データ、及び前記喉頭摘出前に発せられた音声に近似した音声データを格納しても良く、更には全く異なる音質の音声データを格納しても良く、更にまた、喉頭摘出前の音声データを生成し易い音声データを格納しても良い。このように複数の音声データが記憶部24に格納されているとき、音声情報生成部23は、各音声データの関係を例えば関係式等を用いて関連づけを行って選択的に音声データを用いて音声情報を生成しても良い。

【0049】

また、上述の音声生成装置 1 は、サンプリングして記憶部 2 4 に格納した音声データを合成することで音声情報を生成して出力する一例について説明したが、記憶部 2 4 に記憶されている音声データを合成することで生成した音声情報に、音声情報生成部 2 3 により vocoder 処理を施すことにより、サンプリングして記憶されている音声データが示す音声とは異なる音質の音声に変換して出力しても良い。このとき、音声情報生成部 2 3 は、vocoder 処理を用いた例として STRAIGHT を施す。（文献「N. Maeda, et al “Voice Conversion with STRAIGHT,” TECHNICAL REPORT OF IEICE, EA98-9, pp.31-36, 1998」参照）

更にまた、上述の音声生成装置 1 は、マイクロホン 2 1 等からの音声を用いて認識部 2 2 により音声認識処理を行う一例に限らず、例えば使用者及び/又は使用者以外の人が装着する鼻音センサ、呼気流センサ、頸部振動センサからの検出信号及びマイクロホン 2 1 等からの信号を用いて音声認識処理を行っても良い。このように、音声生成装置 1 は、マイクロホン 2 1 のみならず上記各センサを用いることにより、認識部 2 2 による認識率を更に向上させることができる。

【0050】

更にまた、認識部 2 2 は、話者認識 (speaker recognition) 処理を入力される音声について行って各話者に対応した認識結果を生成しても良い。そして、この認識部 2 2 では、各話者に関する情報を認識結果とともにスピーカ部 2 5 や表示部 2 6 に出力することで使用者に提示しても良い。

【0051】

つぎに、この音声生成装置 1 において、音声情報生成部 2 3 により作成した音声情報を出力する機構の種々の例について説明する。なお、本発明は、以下に説明する出力する機構に限られることはないことは勿論である。

【0052】

すなわち、この音声生成装置 1 において、音声情報を出力する機構としてはスピーカ部 2 5 や表示部 2 6 に限らず、例えば骨導や皮膚刺激を利用したものであっても良い。この音声情報を出力する機構は、例えば小さな磁石を鼓膜又は耳小骨に装着し、インダクションコイルを用いて磁石を振動させるものや、側頭部に

ピンを挿入し、電磁的或いは振動子により骨を通して信号を蝸牛に伝達するものであっても良い。

【0053】

このような音声生成装置1は、例えば圧挺板を備え、音声情報生成部23により変換することにより得た信号を前記圧挺板に出力するようにしたものや、皮膚刺激を用いたタクタイルエイド (Tactile Aid) 等の触覚による補償技術を利用したものであっても良く、これらの骨振動や皮膚刺激等を用いた技術を利用することで、音声情報生成部23からの信号を使用者に伝達することができる。皮膚刺激を利用した音声生成装置1においては、音声情報生成部23からの音声情報が入力されるタクタイルエイド用振動子アレイが備えられており、タクタイルエイドと当該振動子アレイを介してスピーカ部25から出力する音声も出力しても良い。

【0054】

また、上述した音声生成装置1の説明においては、音声情報を音声として出力するときの処理の一例について説明したが、これに限らず、例えば人工中耳により使用者に認識結果を提示するものであっても良い。すなわち、この音声生成装置1は、音声情報を電気信号として体外コイルに供給し、体内コイル、振動子を介して使用者に提示しても良い。

【0055】

更には、この音声生成装置1は、人工内耳により使用者に認識結果を提示するものであっても良い。すなわち、この音声生成装置1は、例えば埋め込み電極、スピーチプロセッサ等からなる人工内耳システムに音声情報を電気信号として供給して使用者に提示しても良い。

【0056】

更にまた、この音声生成装置1は、音声情報を超音波帯域の音声に変調・変換して出力してもよく、更には、超音波出力機構 (bone conduction ultrasound) を用いて超音波周波数帯域の信号を生成し、超音波振動子等を介して骨動を通じて使用者に出力しても良い。

【0057】

更にまた、この音声生成装置 1 は、スピーカ部 2 5、表示部 2 6 等の複数の出力手段を備える一例について説明したが、これらの出力手段を組み合わせ用いても良く、更には各出力手段を単独で出力しても良い。

【0058】

なお、上述した音声生成装置 1 の説明においては、マイクロホン 2 1 で検出した音声について音声認識処理、音声生成処理を行う一例について説明したが、使用者等により操作されるキーボード部 2 8 を備え当該キーボード部 2 8 に入力されたデータを音声及び／又は画像とするように認識部 2 2 により変換しても良い。また、このキーボード部 2 8 は、例えば使用者の指に装着され、指の動きを検出することでデータを生成して認識部 2 2 に出力するものであっても良い。

【0059】

また、この音声生成装置 1 は、例えば使用者が液晶画面等をペンにより接触させることで文字及び／又は画像を描き、その軌跡を取り込むことによる画像に基づいて文字及び／又は画像データを生成する文字及び／又は画像データ生成機構を備えていても良い。音声生成装置 1 は、生成した文字及び／又は画像データを認識部 2 2 及び音声情報生成部 2 3 により認識・変換等の処理を行って出力する。

【0060】

更に、この音声生成装置 1 は、例えば自動焦点機能やズーム機能を搭載したデジタルカメラにより動画像や静止画像等を撮像するカメラ機構 2 9 を図 2 に示すように備え、表示部 2 6 に表示するものであっても良い。このカメラ機構 2 9 は例えば図 1 のディスプレイ部 7 と一体に搭載されても良い。また、上記カメラ機構 2 9 としては、デジタルカメラを用いても良い。

【0061】

また、この音声生成装置 1 に備えられたカメラ機構 2 9 は、撮像した画像を使用者の視力や乱視等の状態に合わせて歪ませたり拡大させたりする画像変換処理を施して表示部 2 6 にリアルタイムで表示する眼鏡機能を備えていても良い。

【0062】

このような音声生成装置 1 は、例えばカメラ機構 2 9 から CPU 等からなる信

号処理回路を経由して表示部 26 に撮像した画像を表示する。この音声生成装置 1 は、このようなカメラ機構 29 により例えば話者を撮像した画像を使用者に提示することで、使用者の認識率を向上させる。また、この音声生成装置 1 は、撮像した画像を通信回路 27 を介して外部のネットワークに出力しても良く、更には外部のネットワークからカメラ機構 29 で撮像した画像を入力して通信回路 27 及び信号処理回路等を介して表示部 26 に表示しても良い。

【0063】

更に、この音声生成装置 1 は、静止画撮像用デジタルカメラのようにシャッターを押すことで静止画を撮像しても良い。更に、カメラ機構 29 は、画像とともに、音声を検出して認識部 22 に出力しても良い。このカメラ機構 29 により動画像を撮像するときの信号方式としては、例えば M P E G (Moving Picture Experts Group) 方式などを用いる。更にまた、この音声生成装置 1 に備えられるカメラ機構 29 は、3 次元 (3 D) 画像を撮像することで、話者や話者の口唇を撮像して表示部 26 に表示させることで更に使用者の認識率を向上させることができる。

【0064】

このような音声生成装置 1 は、使用者自身の発した音声や相手の発した音声等及び／又はその場の情景を撮像した画像を記録し再生することで、言語学習における復習することができ言語学習に役立てることができる。

【0065】

また、この音声生成装置 1 によれば、画像を拡大処理等して表示部 26 に表示することで相手を確認し全体の雰囲気をつかめ音声聴取の正確さが向上し、更に読唇 (lip reading) を行うことが可能となり認識率を上昇させる。

【0066】

更にまた、この音声生成装置 1 は、例えばスイッチ機構が設けられており、マイクロホン 21 で検出した音声をスピーカ部 25 により出力するか、カメラ機構 29 により撮像した画像を表示部 26 により出力するか、又は音声及び画像の双方を出力するかを使用者により制御可能としても良い。このときスイッチ機構は、使用者に操作されることで、音声情報生成部 23 から出力を制御する。

【0067】

また例として、スイッチ機構は、使用者及び／又は使用者以外の音声を検出して、例えば「音声」という音声を検出したときにはマイクロホン21で検出した音声をスピーカ部25により出力するように切り換え、例えば「画像」という音声を検出したときにはカメラ機構29により撮像した画等像を表示部26により出力するように切り換え、「音声、画像」という音声を検出したときには音声及び画像の双方を出力するように切り換えても良く、以上のような音声認識を用いたスイッチ制御機構を備えていても良い。

【0068】

更にまた、このスイッチ機構は、カメラ機構29のズーム状態等のパラメータを切り換えることでカメラ機構29で画像を撮像するときの状態を切り換える機能を備えていても良い。

【0069】

更にまた、この音声生成装置1は、スピーカ部25及び／又は表示部26から出力する出力結果を同時に或いは時間差を持たせて出力してするように音声情報生成部23で制御するスイッチ機構を備えていても良く、複数回に亘って出力結果を出力するか一回に限って出力結果を出力するかを制御するスイッチ機構を備えていても良い。

【0070】

また、この音声生成装置1の説明においては、図2に示したような一例について説明したが、入力された音声について上述した種々の変換処理を行って表示部26に表示させる第1の処理を行うCPUと、入力された音声について上述した種々の変換処理を行ってスピーカ部13cに出力結果を出力するための第2の処理を行うCPUと、カメラ機構29で撮像した画像を表示するための第3の処理を行うCPUとを備えたものであっても良い。

【0071】

このような音声生成装置1は、各処理を行うCPUを独立に動作させて第1の処理又は第2の処理を行わせて出力させても良く、更には各処理を行うCPUを同時に動作させて第1の処理、第2の処理、及び第3の処理を行わせて出力させ

ても良く、更には、第1及び第2の処理、第1及び第3の処理又は第2及び第3の処理を行うCPUを同時に動作させて出力させても良い。

【0072】

更にまた、音声生成装置1は、使用者の身体状態、利用状態及び使用目的に応じて上述した種々の出力機構からの出力結果を同時に或いは時間差を持たせて出力するように音声情報生成部23で制御しても良い。

【0073】

更に、この音声生成装置1は、複数のCPUを有し、上述した複数のCPUで行う第1～第3処理のうち、少なくとも1の処理をひとつのCPUで行うとともに、残りの処理を他のCPUで行っても良い。

【0074】

例えば、この音声生成装置1において、ひとつのCPUが入力された音声を文字データとして変換を行って表示部26に出力する処理(text to speech synthesis)を行うとともに、又はひとつのCPUが入力された音声に対して文字データとして変換を行って他のCPUが入力された同じ音声に対してSTRAIGHT処理を行ったりしてスピーカ部25に出力する処理を行い、他のCPUが入力された音声に対してボコーダ処理のうち、例えば音声分析合成法STRAIGHTを用いた処理を行ってスピーカ部25に出力する処理を行っても良い。すなわちこの音声生成装置1は、スピーカ部25に出力する信号と、表示部26に出力信号とで異なる処理を異なるCPUにより行うものであっても良い。

【0075】

更に、この音声生成装置1においては、上述した種々の変換処理を行って上述の種々の出力機構に出力する処理を行うCPUを有するとともに、変換処理を施さないでマイクロホン21に入力された音声を出力しても良い。

【0076】

更に、この音声生成装置1においては、上述した種々の変換処理の一を行うためのCPUと、他の変換処理を行うCPUとを別個に備えていても良い。

【0077】

更に、この音声生成装置1においては、上述のように認識結果や変換した認識

結果や撮像した画像等について音声情報生成部 2 3 で変換する処理を行うとともに、従来の電気喉頭等を用いた代用発声法と同様に音声を検出して得た電気信号を増幅させて音質調整、利得調整や圧縮調整等を行いスピーカ部 1 3 c に出力するものであっても良い。

【0078】

なお、この音声生成装置 1 において、認識部 2 2 及び音声情報生成部 2 3 で行う処理を、例えばフーリエ変換、ボコーダ処理 (STRAIGHT等) の処理を組み合わせることで、上述した処理を行っても良い。

【0079】

また、上述の実施の形態では、本発明を HMD 2 を備えたウェアブルコンピュータに適用した一例について説明したが、所謂フェイスマウントディスプレイに適用しても良い。

【0080】

また、本実施の形態では、HMD 2 と、コンピュータ部 3 との間を光ファイバケーブル 4 で接続してなる携帯型の音声生成装置 1 について説明したが、HMD 2 とコンピュータ部 3 との間をワイヤレスとし、HMD 2 とコンピュータ部 3 との間を無線や赤外線を用いた信号伝送方式等により情報の送受信を行っても良い。更に、この音声生成装置 1 においては、HMD 2 とコンピュータ部 3 との間をワイヤレスとする場合のみならず、図 2 に示した各部が行う機能毎に分割して複数の装置とし、各装置間をワイヤレスとしても良く、コンピュータ部 3 を装着させずに HMD 2 と情報の送受信を行っても良い。更にまた、この音声生成装置 1 においては、使用者の身体状態、利用状態、使用目的に応じて、図 2 に示した各部が行う機能毎に分割して複数の装置とし、各装置間をワイヤレスとしても良い。これにより、音声生成装置 1 は、使用者が装着する装置の重量、体積を軽減し、使用者の身体を向上させ、使用者の認識率を更に向上させることができる。

【0081】

また、本発明を適用した音声生成装置 1 によれば、合成した音声を表示することで使用者に提示することができるので、例えば事務 (ウェアブルコンピュータ

として）、通信（自動翻訳電話への応用など）、産業医学領域（メンタルヘルスなど）、医療現場（聴力検査への利用）、外国語学習、言語訓練、娯楽（テレビゲームやアミューズメントパークでのヴァーチャルリアリティを使用したものなど）、個人用のホームシアター、コンサートや試合等の観戦、番組製作（アニメーション、実写映像、ニュース、音楽制作）、水中（ダイビングでの水中における会話など）、諜報活動や軍事、騒音下などの悪条件での作業業務（建築現場工場など）、スポーツ（自動車やヨット等のレースや、山や海等の冒険時、選手の試合時や練習時での選手同士や選手とコーチ間の意志疎通や情報変換）、や宇宙空間での作業、運輸（宇宙船や飛行機のパイロット）、カーナビゲーションシステム、ヴァーチャルリアリティを用いた種々のシミュレーション作業（遠隔手術など）等にも適用可能であり、音声言語障害者のみならず広い分野で使用可能である。

【0082】

【発明の効果】

以上詳細に説明したように、本発明に係る音声生成装置は、音声言語障害を有しないで発せられた音声を予めサンプリングすることで生成した音声データを記憶する記憶手段と、音声認識手段からの認識結果に基づいて、上記記憶手段に記憶された音声データを用いて出力する音声を示す音声情報を生成する音声情報生成手段とを備えるので、音声言語障害者からの音声を用いて、発声者本来の音声を出力することができ、音声言語障害者等が自然な音声で発声することを可能とする。

【0083】

また、本発明に係る音声生成方法は、音声言語障害を有しないで発した音声をサンプリングし、音声言語障害を有して発せられた音声を検出して音声信号を生成し、音声信号に基づいて音声認識をして認識結果を生成し、認識結果に基づいて、音声言語障害を用いなくて発せられた音声をサンプリングすることで生成した音声データを用いて出力する音声を示す音声情報を生成し、音声情報を用いて音声を出力すること認識結果に基づいて、喉頭摘出後の代用発声法を用いなくて発せられた音声をサンプリングすることで生成した音声データを構成することで

出力する音声を示す音声情報を生成し、音声情報を用いて音声を出力するので、出力する音声の雑音を低減することができるとともに、喉頭摘出した人等が自然な音声で発声することを可能とする。

【0084】

更にまた、本発明に係る他の音声生成方法は、音声認識結果を用いて音声言語障害のないアクセント、イントネーション等が自然な言語を音声を検出された発声者の音声をサンプリングすることで生成した音声データを用いて構成してなる音声情報を生成し、上記音声情報を用いて音声言語障害のない音声を出力することができるので、発声者からの言語が不自然であっても音声言語障害のない音声を出力することができ、難聴等に起因する音声言語障害があっても自然な音声で発声することを可能とさせる。

【図面の簡単な説明】

【図1】

本発明を適用した音声生成装置の外観の一例を示すブロック図である。

【図2】

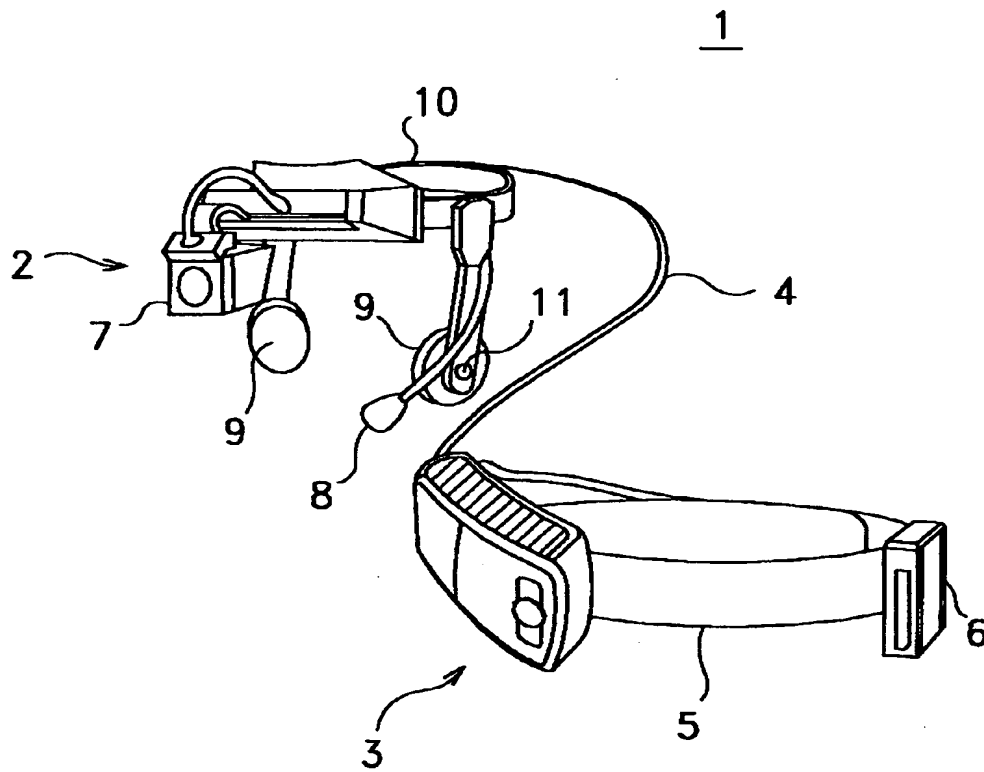
本発明を適用した音声生成装置の構成を示すブロック図である。

【符号の説明】

- 1 音声生成装置、2 ヘッドマウントディスプレイ、3 コンピュータ部、7
- ディスプレイ部、8 使用者用マイクロホン、11 外部用マイクロホン、2
- 1 マイクロホン、23 音声情報生成部、24 記憶部、25 スピーカ部、
- 26 表示部

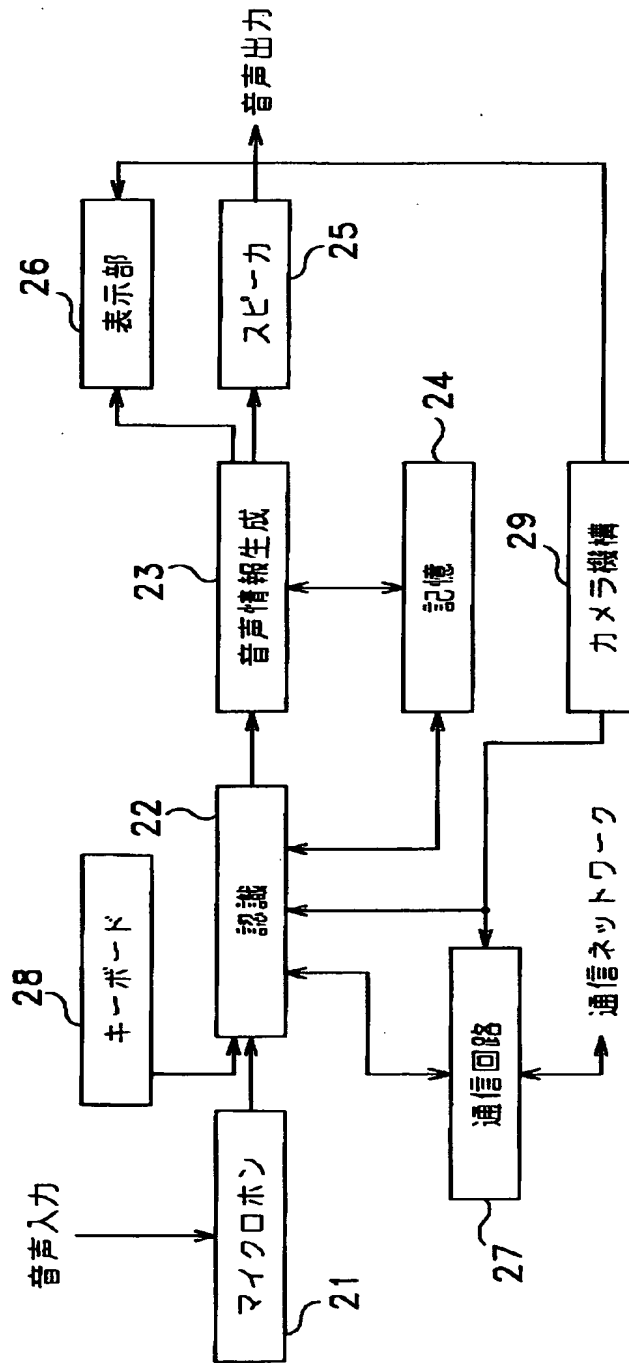
【書類名】 図面

【図 1】



本発明を適用した音声生成装置

【図 2】



本発明を適用した音声生成装置

【書類名】 要約書

【要約】

【課題】 喉頭摘出や舌口腔底切除や構音障害等による音声言語障害を有する人達が本来自身がもつ、或いは自在に変換させて自然な音声で発声させる。

【解決手段】 音声言語障害を有して発せられた音声を検出して音声信号を生成する音響電気変換手段 21 と、音声信号に基づいて音声認識をする処理を行う音声認識手段 22 と、音声言語障害を有しないで発せられた音声を予めサンプリングすることで生成した音声データを記憶する記憶手段 24 と、認識結果に基づいて、上記記憶手段 24 の音声データを用いて出力する音声を示す音声情報を生成する音声情報生成手段 23 と、音声情報を用いて音声を出力する音声出力手段 25, 26 とを備えることで、喉頭を摘出した人やその他の音声言語機能障害を有する人等が本来自身がもつ、或いは自在に変換させて自然な音声で発声することを可能とする。

【選択図】 図 2

出 願 人 履 歴 情 報

識別番号 [399011357]

1. 変更年月日 1999年 2月17日
[変更理由] 新規登録
住 所 東京都港区高輪四丁目11番24号
氏 名 有限会社ジーエムアンドエム

THIS PAGE BLANK (USPTO)